THIS PAGE BLANK (USPTO)

(12) # EUROPEAN PATENT APPLICATION

(72) Inventor : **Raguram, Sasisekharan**
64 Croydon Court
Somerset, New Jersey 08873 (US)
Inventor : **Seshadri, V.**
24 Branson Drive
Lincroft, New Jersey 07738 (US)
Inventor : **Weiss, Sholom M.**
78 Woodbridge Avenue
New Brunswick, New Jersey 08906 (US)

(74) Representative : **Watts, Christopher Malcolm**
Kelway, Dr. et al
AT&T (UK) Ltd.
5, Mornington Road
Woodford Green· Essex, IG8 0TU (GB)

(54) **Data analysis and event· prediction in a telecommunications network.**

(57)   A facility is provided for enhancing an oper-
ations support system so that, based on data
generated as a result of an event occurring in an
associated telecommunications network, the
operations support system can predict the
likelihood of the event occurring again in the
network.

FIG. 1
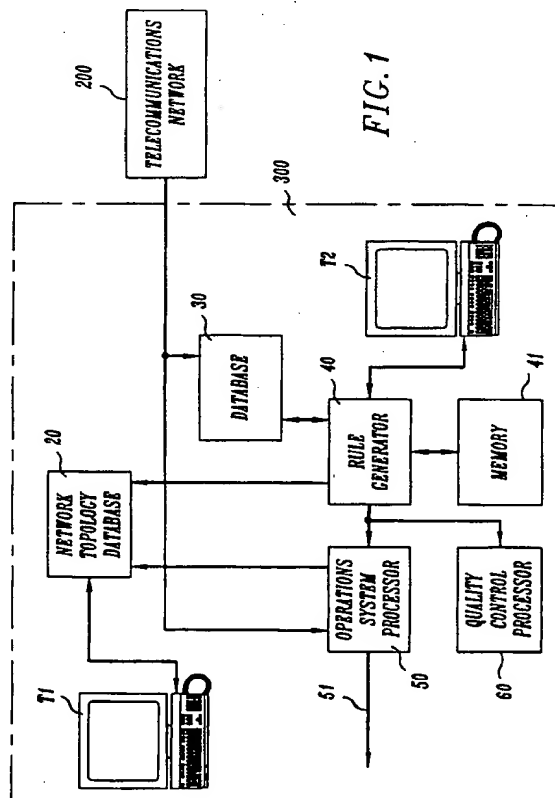
EP 0 650 302 A2

Field of the Invention

The invention relates to processing data generated by a telecommunications network and applying the result of such processing to other network data to predict the possibility of a particular event happening in the network.

Background of the Invention

A large telecommunications network typically generates a large number of reports over a short period of time, e.g., twenty-four hours. Such reports include data relating to, for example, telephone traffic, circuit alarms, etc. A Network Management System (NMS) associated with the telecommunications network accumulates and processes the data to track and manage the performance of the network. Such management includes (a) identifying and repairing network problems when they occur, and (b) trends in telephone traffic patterns. For example, consider the way in which an NMS processes trouble alarms, which may be classified as either transient or non-transient. Non-transient alarms are usually indicative of a "hard" fault. Transient alarms, on the other hand, are usually indicative of either a reoccurring and nonreoccurring network problem. A nonreoccurring alarm is usually the result of an inconsequential fault and, therefore, may be ignored. The reason for this is that a nonreoccurring alarm is most likely the result of some external condition, such as, for example, a strike of lightning, causing a momentary loss of a carrier signal at a point in the network. A reoccurring transient alarm, on the other hand, is usually indicative of a chronic problem, for example, a network facility or circuit that fails when particular conditions occur. Such a problem might affect the quality of telecommunications when it occurs.

In most instances, however, it is difficult to distinguish a chronic alarm from a transient, non-chronic alarm. Moreover, the majority of alarms that are generated in a telecommunications network are of the nonreoccurring type. Because of this, an NMS may spend an inordinate amount of time and resources attempting to identify the sources of respective nonreoccurring alarms, thereby possibly delaying the repair of chronic problems.

Summary of the Invention

An advancement in the art of telecommunications is achieved by arranging an operations support system so that, in accord with an aspect of the invention, it generates a rule set from data generated by an associated telecommunications system and then applies the rule set to data subsequently generated by the telecommunications system to predict that a particular event(s) will likely occur in the system. In an illustrative embodiment of the invention, a set of predetermined measurements is generated for the data that is initially obtained from the telecommunications system and then supplied to a particular rule induction method which then generates a number of rules as a function of such measurements and selects a best rule set from the generated rules such that the selected rule set is capable of predicting accurately the occurrence of past events characterized by the data. The rule set is then used to predict future occurrences of the events.

Brief Description of the Drawing

In the FIGs:
FIG. 1 shows a broad block diagram of an operations system; and
FIG. 2 shows in flow chart form the program which implements the invention in the rule generator of FIG. 1.

Detailed Description

Turning now to FIG. 1 there is shown a broad block diagram of a conventional telecommunications network 200 and operations support system 300. Operations support system 300, more particularly, supports the overall operation of network 200. One aspect of such support deals with tracking the performance of network 200 by accumulating and processing alarm messages generated by individual facilities comprising network 200. An alarm message may be generated by a facility as a result of one of a number of different conditions. Such conditions include (a) a spurious problem that occurs for just a short period of time, (b) a chronic problem that occurs periodically, but which can cause an associated facility to slowly degrade over time, (c) a "hard" failure which may cause the associated facility to become inoperable, etc. The number of such alarms that network 200 may generate within a particular period of time, e.g., 24 hours, could be very high. Such alarms are accumulated by system 300 and stored in database 30. An alarm message includes information identifying a par-

2

ticular problem and the identity of the network 200 facilities that might be affected by the problem.

The alarm messages that are stored in database 30 may be access d by facility type. For example, all alarm messages associated with a particular type of facility, such as echo cancelers, may be accessed by presenting a request identifying the particular type of facility type to database 30. Such access may be refined
5 by also specifying a particular period of time in the request. In response to the request, database 30 outputs an alarm message that occurred over the specified period of time for the identified type of facility.

One dilemma that an operations support system faces is which of the large number of alarms that are generated on a daily basis should be addressed first. It is apparent that alarms that are generated as a result of a "hard" fault are usually easy to identify and are addressed first. However, the majority of alarms that are
10 generated daily are typically due to spurious conditions and chronic problems. Moreover, it is difficult to determine initially whether an alarm occurred as a result of a spurious condition or a chronic problem. As such, much effort is sometimes directed to tracking down the cause of spurious alarms since they comprise the majority of alarms that occur each day. Accordingly, alarms identifying a chronic problem may not be dealt with until the facility generating such alarms degrades to a hard fault.

15 Recognizing that problem, we have adapted an operations support system so that it processes alarms to identify chronic problems and prioritize the repair of those problems. To that end, Operations Support System (OSS) 300 includes rule generator 40 and operations system processor 50 which operate in accordance with the invention to identify and prioritize chronic problems. More particularly, rule generator 40, which may be, for example, embodied in a SPARC station 2 available from Sun Microsystems, Inc., responds to receipt of a
20 command from an external source, e.g., terminal T2, by sending an access request identifying a facility type and period of time to database 30. As mentioned above, database 30, in response to such a request, outputs all alarm messages that occurred over the specified period of time for the type of facility identified in the request, in which the facility type may be, for example, echo cancelers, T1 carriers, T3 carriers, toll switches, etc. The specified period of time is typically one to several weeks and may be identified by a starting date and
25 an ending date.

Assuming that the requested period of time equals two weeks and the facility type is T1 carriers, then th number of alarm messages of the requested type that database 30 outputs could be very large if network 200 employs an appreciable number of T1 carriers. Upon receipt of the alarm messages, generator 40 stores them in its associated memory 41 such that alarm messages associated with a specific facility are stored in a re-
30 spective memory array. When all such messages have been received from database 30 and stored in memory 41, then generator 40 processes the messages and generates, in accord with an aspect of the invention, a set of rules which may be applied to identify which of the alarms stored in memory 41 are indicative of respective chronic problems. The rules are then passed to operations system processor 50, which then applies the rules to alarm messages associated with the same type of facility, but occurring thereafter. Such rules are also
35 passed to processor 60, as discussed below.

The program which controls the operation of processor 40 to generate a particular set of rules for the processing of alarm messages is shown in flow chart form in FIG. 2. Specifically, the program is entered at block 2000 responsive to receipt of the aforementioned command. At block 2000, the program proceeds to block 2001 where it forms a request message containing the identity of the facility type and particular period of time, as
40 discussed above. The program then supplies the message to database 30. As mentioned above, database 30, in turn, (a) unloads the alarm messages from its associated memory for the specified facility type and for the specified period of time and (b) passes them to processor 40. Processor 40, in turn and under control of the program, stores the messages in associated memory 41, as mentioned above. When such alarm messages have been so stored, the program then proceeds to block 2002. (Hereinafter the term "alarm" may also be re-
45 ferred to as an event.)

At block 2002, the program processes the stored messages to obtain a number of different performance measurements for each facility of the specified type. The measurements summarize spatial and temporal information about a facility and associated problems. Such measurements may include the number of (a) different types of alarm messages that occurred for the specific facility, (b) times an alarm (event) occurred during
50 a unit of time for the facility, (c) units of time or (d) the location of the facility, during which the alarm (event) occurred. As will be discussed below, such measurements are determined for first and second windows, W1 and W2, comprising the aforementioned specified period of time. In addition, each window is partitioned into units of time for the purpose of performing such measurements. (An illustrative list of such performance measurements, or features, is shown in Appendix A.)

55 Specifically, blocks 2002, 2003 and 2004 constitute a "do" loop for processing alarms associated with the identified facility. Once it identifies such messages, then the program generates the different performance measurements for each window, W1 and W2, and for the units of time associated with those windows. In addition, the program accesses database 20 to determine the circuit layout for the facility within network 200.

3

The circuit layout includes the identities of the other network 200 elements that may be connected directly to the facility. The program then performs the same processing steps for a next facility of the sp cified type. The program exits the do loop and proceeds to block 2005 upon processing the alarm messages for all of the facilities of the specified type.

At block 2005 the program builds a data file, in which a record of the data file comprises a number of fields for entering respective performance measurements derived for a specific facility and the circuit layout for the facility. The data file also includes a header which identifies the fields forming a record and identifies the contents of the fields. The program proceeds to block 2006 once it has completed the building of the data file. At block 2006, the program enters a rule generation program, which processes the data file using rule induction methods to generate a set of rules for the specified type of facility. For example, a set of rules of the form "if A is > B, then choose class 1", where A may represent a performance measurement, B may represent a particular value and class 1 may represent a chronic problem.

(It is noted that the rule induction program represented by block 2006 follows the rule induction technique disclosed in an article entitled "Reduced Complexity Rule Induction", by S. Weiss and N. Indurkhya and published in the Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), Sydney, Australia, 1991, pages 678-684, which is hereby incorporated by reference. Alternatively, a number of rule induction programs are commercially available, such as the C4.5 program available from Morgan Kaufmann Publishers, San Mateo, California, U.S.A.).

Briefly, rule induction methods attempt to find a compact rule set that completely covers and separates various classes of data, e.g. alarm messages respectively relating to spurious and chronic problems. A covering set of rules is found by heuristically searching for a single best rule that covers different cases, e.g., records in the data file, for only one class. Once a best conjunctive rule is found for a particular class e.g., chronic or nonchronic problems, the rule is then added to the rule set and the cases satisfying it are removed from further consideration. This process is repeated until there are no remaining cases to be covered. Once a covering rule set is found that separates the classes, the induced set of rules is further refined by using either pruning or statistical techniques. The initial covering rule set is then scaled back to the most statistically accurate subset of rules using training and testing evaluation methods.

When the program has generated a number of sets of rules, it applies those rule sets to the test data portion of the data file generated at block 2005. In doing so, it prunes the rule sets so that an error rate defining the accuracy of each rule set may be minimized when applied to the test data. That is, the program strives to improve the accuracy of a rule set to predict the occurrence of chronic as well as nonchronic alarms. Following the foregoing, the program then proceeds to block 2007.

The program at block 2007 selects a rule set from the sets of rules generated at block 2006, in which the selection is based on a specification provided by an external source, for example, a user or another program. Such a specification may request a rule set having a particular predictive accuracy for a certain class of problems, such as chronic problems. The specification may also include coverage for a particular class of problems. Coverage as defined herein is a measure of the range of particular problems, e.g., chronic problems, that a rule set can predict accurately. For example, the following illustrates one possible rule set:

$ALRD > x_1$
$FBL > x_2$
$SLL > x_3$
$OFL > x_4$ or
$BVL > x_5$

where x_1 through x_5 are particular values that are generated by the rule generator. The coverage of the above rule set may be increased by adding, for example, the following rule to form a rule set of 2.

$CT2 \& ALRCT > x_6 \& ALRH > x_7;$

where x_6 and x_7 are also values generated by the rule generator.

It is noted, however, that increasing the coverage of a rule set may lead to a decrease in the predictive accuracy of the rule set. The reason for this is that as the number of cases that are covered increases it is likely that errors will occur.

When the program at block 2007 selects a rule set in accordance with the externally provided specification, it then proceeds to block 2008, where it outputs the rule set to processor 50 and processor 60. The program then proceeds to block 2009 where it checks to see it is done, that is, if the program has been directed to generate a rule set for another type of facility. If that is the case, then the program proceeds to block 2010 where it obtains the identity for the specified facility and then returns to block 2001. Otherwise, the program exits.

Returning to FIG. 1, operations system processor 50 applies the rule set to new alarm messages that it obtains from communications network 200 so that it can accurately predict the occurrence of a particular chronic problem and output a message to that effect via path 51. In this way, processor 50 may direct attention away

from spurious problems (nonchronic problems) and focus attention on chronic problems so that such problems may be dealt with before they degrade into a hard fault.

The rule set outputted by processor 40 may also be used to predict trends in the occurrence of chronic problems. Quality control processor 60 uses the rule set to determine such trends, and uses the trends in con-
5   junction with other quality control criteria, e.g., the rate at which chronic problems are being repaired, to eval-
uate the performance of network 200. For example, if a trend indicates that a particular section of network 200 may exceed its number of allowable chronic problems for given peiriod of time, then processor 60 uses that trend to output a performance value for that section of network 200.

As mentioned above, rule induction can be applied to other applications as a way of generating rule sets
10   for those applications. For example, one application may be traffic data in which rule induction may be applied to such data to generate a set of rules that may be used to predict particular trends in future traffic patterns. In particular, traffic data may be collected over period of time and stored in a database, e.g., database 30. The above-described program at block 2001 (FIG. 2) may then access that data as well as data indicative of the topology of the network in the manner discussed above. The program may then divide the traffic data into re-
15   spective windows W1 and W2, and then generate a rule set for the data. The measurements made in W1 are used to predict the measurements in W2. If the prediction is accurate, then the rule set may be used to predict future traffic patterns, in which such patterns may be, for example, predicting appropriate levels of blocking for particular telephone routes.

As another example, when a disaster strikes a particular area, e.g., a hurricane, particular controls are
20   activated to automatically block low priority calls so that such calls do not impact the processing and routing of high priority calls. The levels of such controls can be predicted accurately by processing traffic data accu-
mulated during a previous disaster, in accord with the invention, to generate a rule set that can be used to predict such levels of control.

The foregoing is merely illustrative of the principles of the invention. Those skilled in the art will be able
25   to devise numerous arrangements, which, although not explicitly shown or described herein, nevertheless em-
body those principles that are within the spirit and scope of the invention.

30

APPENDIX A

Some examples of Performance Measurements for facilities with failures:

35

Measurement Name	Description
ALRD	Number of time units with alarms
ALRH	Number of time units with alarms during the last day
ERHD	Number of time units with exceptions over a threshold H
ERLD	Number of time units with exceptions over a threshold L
OFH	Number of time units with out-of-frame exceptions over a threshold H
OFL	Number of time units with out-of-frame exceptions over a threshold L
FBL	Number of time units with framing bit exceptions over a threshold L
BVL	Number of time units with Bi-polar violations over a threshold L
SLL	Number of time units with slips over a threshold L
ALRCT	Total number of alarms
AVGAL	Average number of alarms per time unit with alarm
OST	Seconds facility was out of service
CT1	Type of circuit - Domestic circuit
CT2	Type of circuit - International circuit
CT3	Type of circuit - Private Lines.

5

Claims

1. A method of processing particular data to generate at least one rule set operable for predicting particular events that may occur in a telecommunications system, said particular data being data generated by said telecommunications system and includes data characterizing prior ones of said events, said method comprising the steps of

obtaining said particular data from said telecommunications system, and CHARACTERIZED BY generating predetermined measurements for the particular data and generating at least one rule set as a function of said measurements such that said prior ones as well as future ones of said events may be predicted using said at least one rule set.

2. The method set forth in claim 1 FURTHER CHARACTERIZED BY the step of applying said at least one rule set to data currently generated by said telecommunications system to predict future occurrences of said events.

3. The method set forth in claim 1 FURTHER CHARACTERIZED IN THAT said step of generating includes the step of generating a plurality of rule sets and selecting as said at least one rule set that one of said plurality of rule sets which meets a predetermined specification.

4. The method set forth in claim 3 FURTHER CHARACTERIZED IN THAT said predetermined specification includes a predetermined level of accuracy and a predetermined level of coverage.

5. The method set forth in claim 3 FURTHER CHARACTERIZED IN THAT said step of generating includes the steps of

generating a data file comprising said predetermined measurements, and
supplying said data file to a rule induction generator to generate said plurality of rule sets.

6. The method set forth in claim 1 FURTHER CHARACTERIZED IN THAT said particular data defines respective chronic problems that occurred in said telecommunications system.

7. The method set forth in claim 1 FURTHER CHARACTERIZED IN THAT said particular data defines different traffic patterns within said telecommunications system.

8. The method set forth in claim 1 FURTHER CHARACTERIZED IN THAT said step of generating includes the step of

supplying said at least one rule set to an operations system processor so that said processor may use said at least one rule set to predict the occurrence of said events in said system.

9. The method set forth in claim 1 FURTHER CHARACTERIZED IN THAT said step of generating includes the step of

supplying said at least one rule set to a quality control processor so that said quality control processor may predict trends in the occurrence of said events.

10. The method set forth in claim 1 FURTHER CHARACTERIZED IN THAT said particular data is service provisioning information associated with a particular service feature provided by said telecommunications system.
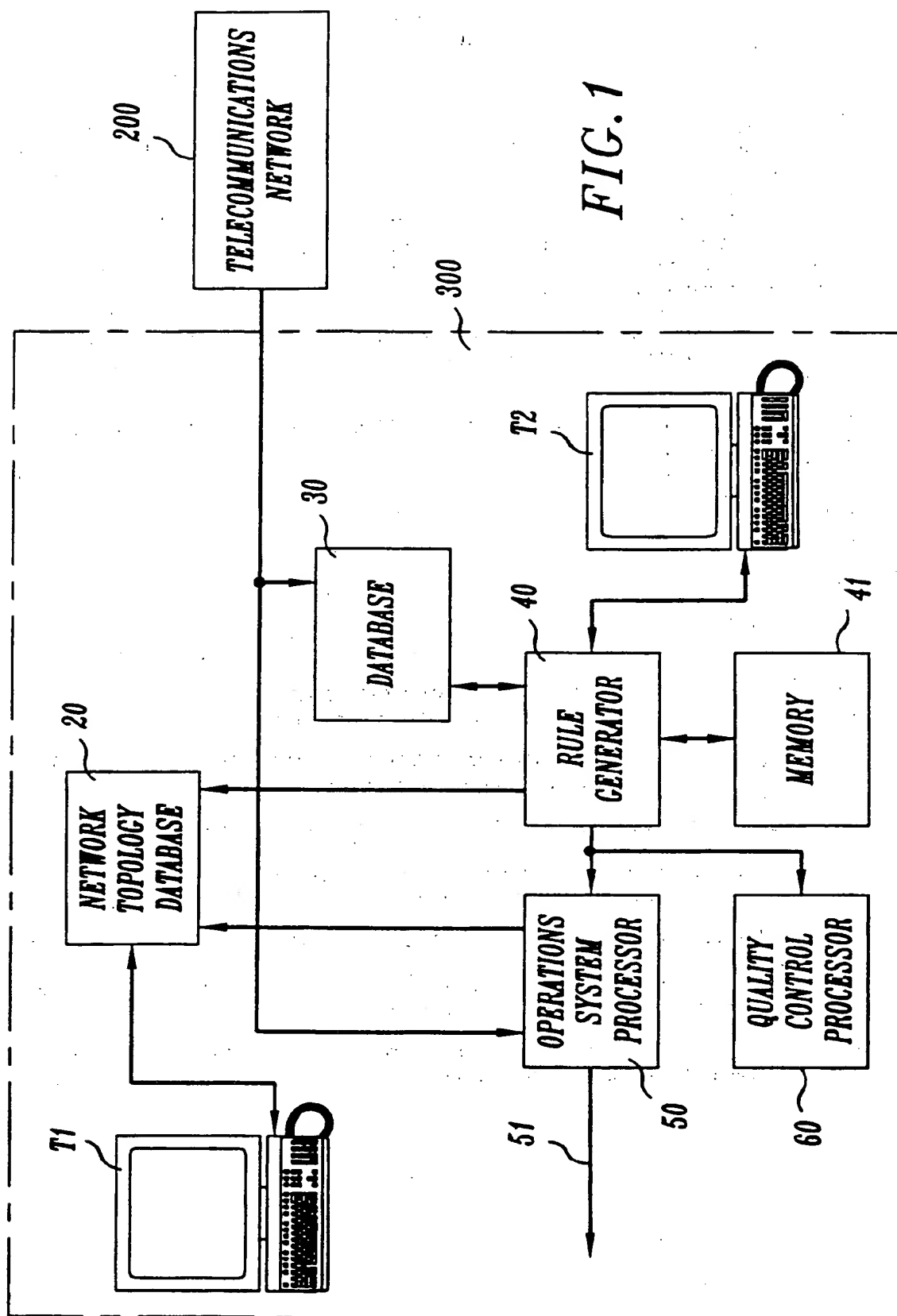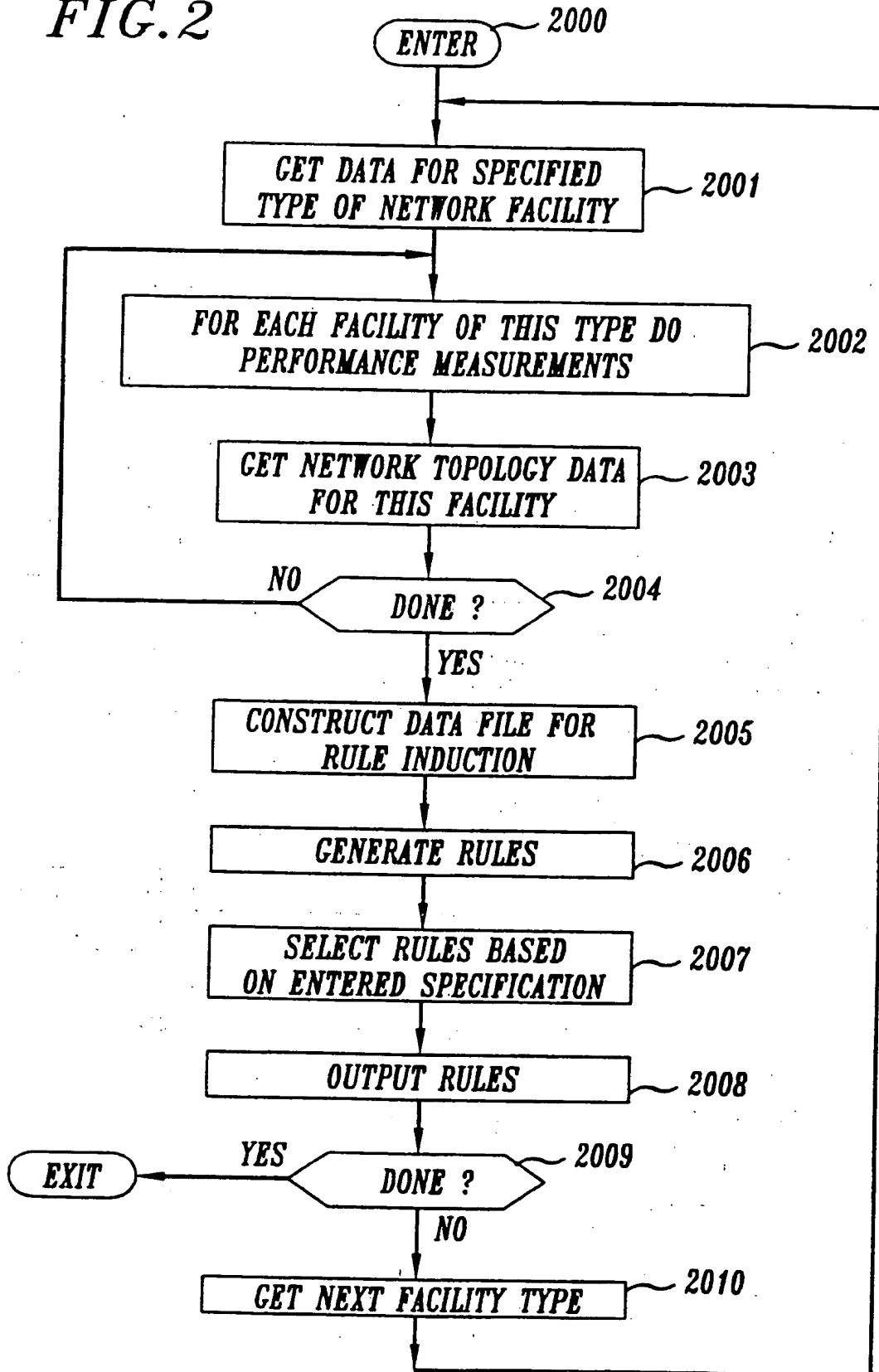
FIG. 1

FIG.2

ENTER — 2000

GET DATA FOR SPECIFIED
TYPE OF NETWORK FACILITY — 2001

FOR EACH FACILITY OF THIS TYPE DO
PERFORMANCE MEASUREMENTS — 2002

GET NETWORK TOPOLOGY DATA
FOR THIS FACILITY — 2003

NO ◁ DONE ? ▷ — 2004

YES

CONSTRUCT DATA FILE FOR
RULE INDUCTION — 2005

GENERATE RULES — 2006

SELECT RULES BASED
ON ENTERED SPECIFICATION — 2007

OUTPUT RULES — 2008

EXIT ◁ YES DONE ? ▷ — 2009

NO

GET NEXT FACILITY TYPE — 2010

(72) Inventors:
   • Raguram, Sasisekharan
     Somerset, New Jersey 08873 (US)
   • Seshadri, V.
     Lincroft, New Jersey 07738 (US)
   • Weiss, Sholom M.
     New Brunswick, New Jersey 08906 (US)

(74) Representative:
   Watts, Christopher Malcolm Kelway, Dr. et al
   Lucent Technologies (UK) Ltd,
   5 Mornington Road
   Woodford Green Essex, IG8 0TU (GB)

(54)   **Data analysis and event prediction in a telecommunications network**

(57)   A facility is provided for enhancing an operations support system so that, based on data generated as a result of an event occurring in an associated telecommunications network, the operations support system can predict the likelihood of the event occurring again in the network.
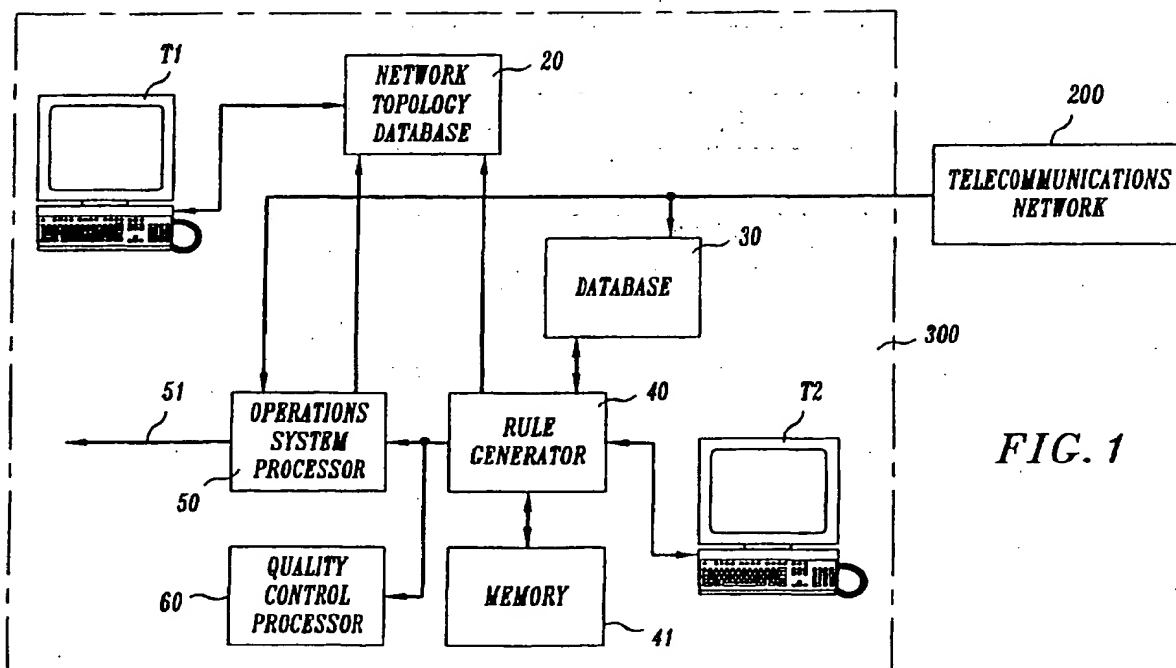
FIG. 1

European Patent Office

EUROPEAN SEARCH REPORT

Application Number

EP 94 30 7460

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
X	GLOBECOM 92, 6 - 9 December 1992, ORLANDO, US, pages 560-564, XP000357845 YING-DAR LIN ET AL: "A FRAMEWORK FOR LEARNING AND INFERENCE IN NETWORK MANAGEMENT" * figure 2 * * page 561, column 1, paragraph 2 - page 562, column 2, paragraph 1 * * page 563, column 2, paragraph 3 - page 564, column 1 *	1-3,5, 7-9	H04Q3/00
X	INTEGRATED NETWORK MANAGEMENT, II, PROC. 2ND INT. SYMPOSIUM, 1 - 5 April 1991, AMSTERDAM, NL, pages 541-549, XP000444253 GOODMAN R M ET AL: "AUTOMATED KNOWLEDGE ACQUISITION FROM NETWORK MANAGEMENT DATABASES" * page 541, paragraph 1 - page 542, paragraph 2 * * page 543, paragraph 3 - page 544, paragraph 2 * * page 545, paragraph 2 * * page 546, paragraph 1 *	1-3,5,8, 9	
A	SUPERCOMM/ICC 92, 14 June 1992, CHICAGO, US, pages 1563-1567, XP000337968 BISIO R ET AL: "USING A MACHINE LEARNING TOOL IN DIAGNOSIS OF NETWORK OVERLOAD" * 1 Introduction * * figure 1 * * 2 Problem description * * 4 MERLINO Machine Learning Tool * * 6 Conclusions *	1-3,5,7, 9	TECHNICAL FIELDS SEARCHED (Int.Cl.6) H04Q H04M

-/--

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	29 July 1997	Neves Appelt, D

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

European Patent Office

EUROPEAN SEARCH REPORT

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
P,X	GLOBECOM '93, 29 November 1993 - 2 December 1993, HOUSTON, US, pages 217-222, XP000428057 SASISEKHARAN R ET AL: "PROACTIVE NETWORK MAINTENANCE USING MACHINE LEARNING" * the whole document *	1-10	
A	NOMS '92, 6 - 9 April 1992, MEMPHIS, US, pages 520-532, XP000344720 STINSON W ET AL: "A STATE-BASED APPROACH TO REAL-TIME TELECOMMUNICATIONS NETWORK MANAGEMENT" * the whole document *		
			TECHNICAL FIELDS SEARCHED (Int.Cl.6)

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	29 July 1997	Neves Appelt, D

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

THIS PAGE BLANK (USPTO)